Introduzione

Nella panoramica della stampa britannica, *broadsheet* e *tabloid* rappresentano due modelli editoriali distinti, non solo per il formato, ma anche per contenuti, stile e pubblico di riferimento. Negli ultimi anni, tuttavia, diversi studi hanno messo in luce la progressiva ascesa del modello *tabloid* e parallelamente, l'affermarsi del fenomeno noto come *tabloidization*[1]. Tale processo avrebbe innescato cambiamenti significativi sia nello stile che nei contenuti dei giornali. In particolare, sul piano stilistico, si osserva una transazione da una scrittura analitica e complessa verso una comunicazione più colloquiale e sensazionalistica, contraddistinta da frasi brevi e da una maggiore enfasi sulla dimensione personale.

La tendenza riscontrata ha suscitato l'interesse di numerosi studi linguistici e mediatici, in quanto riflette non soltanto cambiamenti editoriali, ma anche trasformazioni sociolinguistiche di più ampia portata. Il presente studio si propone di analizzare le **differenze stilistiche tra *broadsheet* e *tabloid***, adottando l'approccio multidimensionale di Biber[2] e il metodo di raggruppamento per *clusters*, con l'obiettivo di individuare le strategie linguistiche impiegate nei due sottogeneri giornalistici e verificare l'esistenza di una distinzione significativa nei rispettivi stili. In tal senso, la ricerca mira a esplorare la **variazione stilistica all'interno del registro *press editorials***, individuato da Biber, al quale appartengono entrambe le tipologie di giornale. Ai fini di un'interpretazione accurata, è importante sottolineare che si tratta di una comparazione stilistica interna a uno stesso registro; pertanto, anche differenze di entità limitata verranno considerate potenzialmente significative.

Per l'analisi che segue sono state considerate cinque delle sei Dimensioni individuate da Biber. La sesta Dimensione è stata esclusa per diverse ragioni. In primo luogo, numerosi studi precedenti ne hanno evidenziato la limitata rilevanza empirica, osservando come essa contribuisca poco alla distinzione tra generi testuali. Inoltre, la Dimensione 6 si fonda su un numero esiguo di variabili linguistiche e presenta una coerenza interna debole, fattori che ne riducono significativamente l'utilità analitica.

[1] 'Tabloidization.' *Communication*, iResearchNet, https://communication.iresearchnet.com/media/tabloidization/
[2] Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Metodologia

La presente ricerca si compone di tre fasi: raccolta dei dati, analisi e interpretazione dei risultati.

Raccolta dei dati

Nella prima fase, i testi sono stati reperiti tramite la piattaforma *Lexis Nexis* e successivamente raccolti e organizzati in un corpus tramite l'utilizzo del software *Sketch Engine*; entrambe le piattaforme sono accessibili online. Il corpus è costituito da articoli in lingua inglese relativi alle elezioni per la leadership del Partito Conservatore britannico, tenutesi nel 2024. L'accesso alla piattaforma *Lexis Nexis* è stato reso possibile tramite una VPN autorizzata dall'Università di Catania. La ricerca è stata effettuata applicando specifici filtri, tra cui: selezione della lingua (inglese), definizione di intervalli temporali e utilizzo dei nomi dei candidati come parole chiave.

Successivamente, i testi selezionati sono stati codificati in UTF-8 e salvati in formato .txt. È seguita una fase di pulizia eseguita mediante *Sublime Text*, durante la quale sono stati eliminati i duplicati e gli articoli non pertinenti. In questa fase è stata inoltre effettuata una classificazione dei documenti mediante l'inserimento di metadati, in modo da renderli compatibili con le piattaforme utilizzate per l'analisi. Tra questi, sono stati essenziali per lo studio i tag <doc>, volti a identificare ogni singolo articolo, e <docx>, per l'identificazione della testata giornalistica.

Il **corpus** così costituito, denominato *CP Leadership Election*, comprende 417.180 token, 497 articoli provenienti da 10 testate giornalistiche, equamente suddivise tra *broadsheet* e *tabloid*. I quotidiani inclusi nella categoria *broadsheet* sono: *The Times*, *The Guardian*, *The Independent*, *i-News* e *The Daily Telegraph*; per i *tabloid* sono stati selezionati: *Daily Express*, *Daily Mail*, *Daily Mirror* e *The Sun*, *Daily Star*. Si tratta dunque di un corpus specialistico monolingua, in quanto focalizzato sull'ambito del giornalismo politico inglese. Per assicurare un'analisi il più possibile esaustiva e rappresentativa, gli articoli sono stati raccolti integralmente.

Ai fini della ricerca, il corpus è stato suddiviso in due subcorpora distinti, corrispondenti alle due tipologie giornalistiche analizzate. Il subcorpus dei *broadsheet* comprende 315.636 token e 273.471 parole, rappresentando circa il 75% del corpus totale. Quello dei *tabloid*, invece, si compone di 100.557 token e 87.123 parole, corrispondente al restante 25%. Un simile sbilanciamento è da attribuire alle caratteristiche intrinseche di ciascun gruppo. È opportuno sottolineare che alcune testate hanno un peso maggiore all'interno dei rispettivi subcorpora: ad esempio, *Daily Express* rappresenta il 67% del subcorpus dei *tabloid*, mentre *The Guardian* e *The Independent* costituiscono oltre il 30% ciascuno di quello dei *broadsheet*.

Tabella 1

GRUPPO	TOKENS	%

BROADSHEET	315,636	75.659
TABLOID	100,557	24.104

Per rispettare le assunzioni alla base dell'analisi, ogni testata è stata considerata come un'entità autonoma, dotata di uno stile redazionale proprio. Inoltre, per poter utilizzare tutte le 67 variabili linguistiche definite da Biber, è stato selezionato un numero di testi almeno cinque volte superiore al numero delle variabili.

Analisi multidimensionale

Per condurre l'analisi multidimensionale è stato utilizzato il programma sviluppato da Andrea Nini, *MAT – Multidimensional Analysis Tagger*-[3], che replica l'approccio multidimensionale proposto da Biber. L'affidabilità di *MAT* è stata validata tramite analisi condotte su LOB e Brown Corpus, i cui i risultati sono stati confrontati con quelli ottenuti da Biber.

Il programma si avvale del sistema *Stanford* appositamente implementato per il *tagging* delle variabili linguistiche[4]; per maggiori informazioni sul sistema *Stanford* si veda il sito https://nlp.stanford.edu/software/tagger.shtml. Inoltre, al fine di riprodurre in modo più accurato possibile l'analisi di Biber, il programma impiega l'algoritmo definito dallo studioso per il calcolo dei valori statistici. Considerati tali aspetti, non sono stati impiegati ulteriori strumenti statistici per calcolare la significatività statistica dei risultati ottenuti.

Per effettuare l'analisi sono stati considerati i seguenti parametri: rapporto *type/token* con valore 400, per compensare la diversa lunghezza dei testi; nessuna correzione degli *z-score*; utilizzo esclusivo dei *VASW tags*, ossia i tag individuati da Biber.

L'output generato da *MAT* comprende vari dataset: (I) i testi annotati con le variabili linguistiche individuate da Biber; (II) le frequenze relative (per 100 token) delle 67 variabili per ciascun testo; (III) i *dimension scores*, ovvero i punteggi ottenuti da ciascun testo per ciascuna Dimensione. I valori delle frequenze relative sono stati normalizzati su base 10.000 e successivamente aggregati per calcolare le medie relative alle variabili linguistiche e alle dimensioni, per ciascun gruppo. I dataset contenenti i valori statistici sono riportati in Appendice.

Analisi dei cluster e verifiche contrastive

I dati estratti da *MAT* sono stati manipolati al fine di costruire il dataset destinato all'analisi dei clusters, effettuata mediante l'uso di *Lancaster Stats Tools*. Per la creazione del dataset sono stati

[3] Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), *Multi-Dimensional Analysis: Research Methods and Current Issues*, 67-94, London; New York: Bloomsbury Academic.
[4] Ibidem.

assegnati dei *registri* fittizi alle testate giornalistiche, in modo da poterli distinguere chiaramente sul dendogramma. Per la generazione di quest'ultimo sono stati adottati i seguenti parametri: distanza *Manhattan*, nessuna normalizzazione degli *z-score* e metodo *Ward* per l'aggregazione gerarchica.

Al fine di verificare e approfondire la coerenza delle tendenze osservate, sono state condotte analisi contrastive supplementari utilizzando la piattaforma *Sketch Engine*. L'adozione di un secondo strumento analitico ha permesso, pertanto, di effettuare verifiche incrociate, contribuendo a garantire la solidità e l'affidabilità dei risultati ottenuti.

Per l'analisi della Dimensione 1, sono state generate, per entrambi i subcorpora, le liste delle parti del discorso più frequenti, definite *POS*. È opportuno sottolineare la differenza tra *POS* e *POS tag*: i primi forniscono informazioni esclusivamente riguardo alla parte del discorso rappresentata dal token, i secondi includono informazioni morfosintattiche più dettagliate[5]. Per la consultazione completa dei sistemi di annotazione adottati dal software, si rimanda alla sezione dedicata sul sito: *Glossary – English PoS Tagset (CP Leadership Election)*, Sketch Engine, https://www.sketchengine.eu/glossary/pos/.

Relativamente alla Dimensione 2, sono state prodotte le liste dei verbi più frequenti, sulla base dell'elenco di verbi dichiarativi proposto da Biber. Inoltre, mediante lo strumento *Concordance* e l'utilizzo del linguaggio *CQL (Corpus Query Language)*, sono state calcolate le frequenze dei verbi al passato.

La medesima procedura è stata adottata per la Dimensione 3, al fine di calcolare le frequenze relative delle variabili linguistiche esaminate, ovvero la coordinazione frasale e la nominalizzazione. Per condurre l'analisi quantitativa degli avverbi di tempo e di luogo, per l'individuazione dei quali è stata seguita la classificazione proposta da Biber[6], sono state generate liste ordinate per frequenza.

Per la Dimensione 4, è stato riadottato lo strumento *Concordance*. In particolare, sono state generate tre espressioni regolari per il linguaggio *CQL*, per misurare le frequenze di verbi modali, proposizioni infinitive e verbi di persuasione.

Per la Dimensione 5, infine, *Sketch Engine* è stato impiegato per indagare la presenza di strutture linguistiche complesse all'interno dei testi. A tal fine, sono state formulate tre espressioni regolari per il linguaggio *CQL*, con l'obiettivo di quantificare le occorrenze di tre fenomeni linguistici: costruzioni passive introdotte da *by*, avverbi subordinativi e connettivi logici.

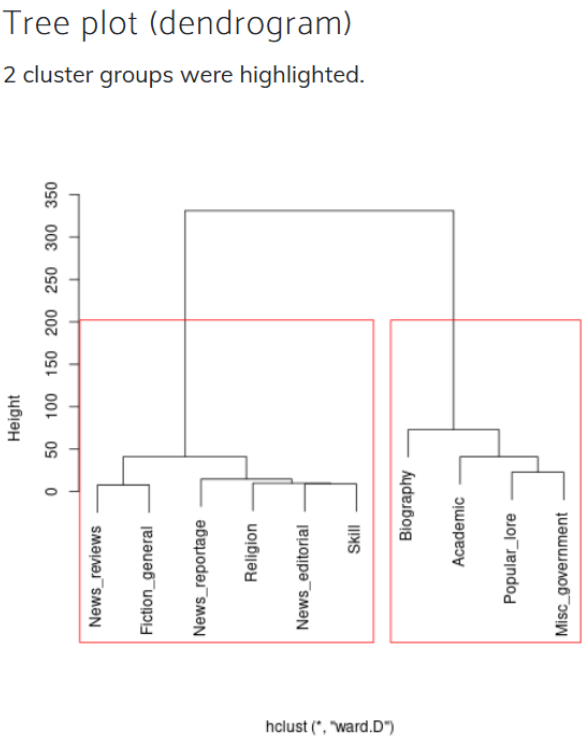
[5]Per l'elenco completo dei POS tag adottati si veda https://www.sketchengine.eu/english-treetagger-pipeline-2.
[6] Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press, Appendice II.

Analisi

Hierarchical Agglomerative Cluster Analysis

Per un'esplorazione visuale del corpus forniamo il risultato della *Hierarchical Agglomerative Cluster Analysis* ottenuta dagli strumenti di *Lancaster Stats Tools*. Il dataset è stato ricavato dall'output di *MAT* : questo è stato diviso per giornali, attribuendo a ciascuno un registro fittizio e calcolando le medie delle frequenze relative su base 10,000 per tutte e 67 le variabili linguistiche e tutti e 10 i giornali. Per la nostra analisi abbiamo adoperato la *Manhattan Distance*, in ragione della sua maggiore efficacia in presenza di *outliers*, e il *Ward's Method*, per la sua capacità di generare cluster compatti e facilmente interpretabili, non è stata effettuata correzione dei dati in *z-scores*. Infine, abbiamo chiesto al software di evidenziare due gruppi.

Figura 1



Osservando il **dendogramma** in *Figura 1*, è possibile vedere in maniera evidente una scissione tra *broadsheet* - gruppo a sinistra - e tabloid - gruppo a destra -, fatta eccezione per il *Daily Express* (*Fiction_general*) che fa cluster con i primi pur essendo in linea teorica un giornale appartenente alla sottocategoria dei *tabloid*. Questo dato non è da trascurare, costituendo il giornale in questione il 67% del subcorpus *tabloid*. In linea generale osserviamo che le distanze tra i due gruppi sono notevoli, soprattutto se comparate alla lunghezza, considerevolmente minore, dei bracci che collegano tra loro i giornali di uno stesso gruppo. Per la legenda riportante la corrispondenza tra registro fittizio e testata consultare "Appendice, Dataset per Cluster Analysis".

Analisi Dimensione 1

La prima Dimensione individuata da Biber[7] distingue testi 'informativi', collocati sul semiasse negativo verso l'estremo *Informational*, e testi 'coinvolti', tendenti verso l'estremo positivo *Involved*. Mentre i primi sono centrati sull'informazione oggettiva e sul contenuto, i secondi sono caratterizzati da un tono personale e prediligono l'espressione soggettiva.

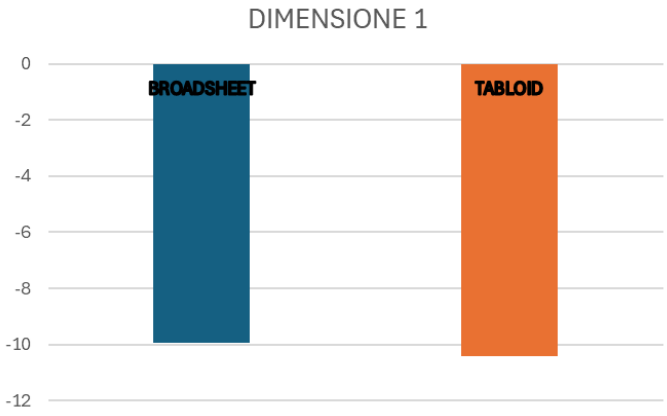
Confrontando le medie dei *dimension scores* dei gruppi analizzati, si evidenzia una tendenza, seppur contenuta, dei *tabloid* a presentare contenuti più orientati all'informazione, collocandosi più in basso, rispetto ai *broadsheet*, sul semiasse negativo.

Come si evince dalla Tabella 2, semplificata dal grafico sottostante, lo scarto tra le due medie, pari a 0.51, indica la presenza di differenza stilistica, anche se non particolarmente marcata.

Tabella 2

GRUPPO	MEDIA DIMENSION SCORE	VALORE MINIMO	VALORE MASSIMO	STANDARD DEVIATION
BROADSHEET	-9.92	-22.15	12.75	6.653936043
TABLOID	-10.43	-27.37	22.95	7.896025982

Figura 2



La discrepanza osservata nelle medie dei due gruppi è stata ulteriormente analizzata attraverso il confronto delle frequenze relative (FR) delle variabili linguistiche che definiscono l'aspetto 'informativo' della lingua[8].

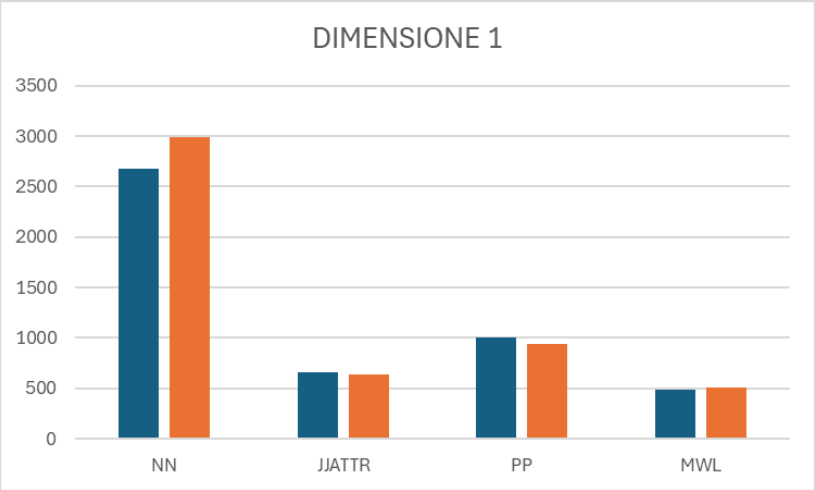
I valori delle FR sono presentati nella Tabella 3, seguiti dal grafico (*Figura 3*) che ne facilita l'interpretazione.

[7] Ivi, pp. 107-108.
[8] Ivi, pp. 102-104.

Tabella 3

GRUPPO	NN	JJATTR	PP	MWL
BROADSHEET	2675.582492	655.3569024	1001.973064	485.3569
TABLOID	2992.08	632.92	940.435	503.27

Figura 3



L'analisi comparativa ha rivelato differenze significative tra i due subcorpora nelle frequenze relative di **nomi** (NN) e **sintagmi preposizionali** (PP), confermando i risultati precedentemente ottenuti. Al contrario, gli **aggettivi in funzione attributiva** (JJATTR) e la lunghezza media delle parole (MWL) mostrano un minore scarto e una tendenza invertita che contraddice l'esito dell'analisi.

Per approfondire ulteriormente sul contrasto rilevato, sono state esaminate le liste delle parti del discorso più frequenti per ciascun subcorpus, ottenute tramite *Sketch Engine*.

Il confronto tra le due liste ha messo in luce alcune differenze, seppure lievi, nel posizionamento e nelle frequenze delle variabili linguistiche oggetto di interesse, ovvero nomi (n), aggettivi in funzione attributiva (j) e preposizioni (i).

Tabella 4

subcorpus: TABLOID	Item	Frequenza	Frequenza Relativa
1.	n	26965	268156.369
2.	x	25852	257088.0197
3.	v	16309	162186.6205
4.	i	10862	108018.3379
5.	j	6626	65892.97612
6.	d	5895	58623.46729
7.	a	4062	40394.99985

8.	c	2552	25378.64097
9.	m	1434	14260.56863

Tabella 5

subcorpus: BROADSHEET	Item	Frequenza	Frequenza Relativa
1)	x	83372	264139.7052
2)	n	79230	251016.9943
3)	v	50908	161287.0522
4)	i	36038	114175.8228
5)	j	21283	67428.93713
6)	d	18111	57379.38638
7)	a	13960	44228.16155
8)	c	8068	25561.08936
9)	m	4666	14782.85113

Nello specifico, i nomi figurano al primo posto nella classifica del subcorpus dei *broadsheet*, mentre si collocano al secondo posto nei *tabloid*, dove si rileva una frequenza relativa leggermente inferiore. Diversamente aggettivi attributivi e preposizioni si posizionano allo stesso posto e mostrano un numero maggiore di frequenze relative all'interno del gruppo *broadsheet*. I confronti effettuati confermano le discrepanze riscontrate nell'analisi precedente.

Questi risultati contrastanti potrebbero essere attribuiti a diversi fattori. In primo luogo, la differenza sostenuta tra le medie dei gruppi suggerisce una sostanziale somiglianza tra i tipi di giornale rispetto alla funzione comunicativa esaminata. Oltre a ciò, lo sbilanciamento del corpus e in particolare, il peso del *Daily Express* – si veda "Introduzione" - il cui stile meriterebbe uno studio approfondito, potrebbe aver inciso nel determinare i contrasti rilevati. Infine, l'elevata variabilità interna tra i due gruppi, evidenziata da un ampio range e da una marcata distanza dalla media (*standard deviation*), potrebbe aver attenuato i risultati.

Complessivamente, le analisi effettuate mostrano una considerevole tendenza, di entrambi i gruppi, verso l'estremo Informational, confermando il carattere orientato all'informazione oggettiva tipico delle testate giornalistiche.

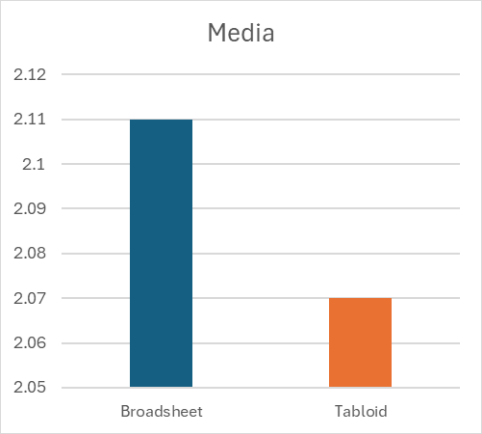
Analisi Dimensione 2

La Dimensione 2 individuata da D. Biber si caratterizza per l'opposizione tra testi ad alto contenuto narrativo e testi a carattere prevalentemente espositivo o informativo. Un elevato *dimension score*

segnala la presenza di tratti narrativi marcati, mentre un valore basso denota un'impostazione più non-narrativa.

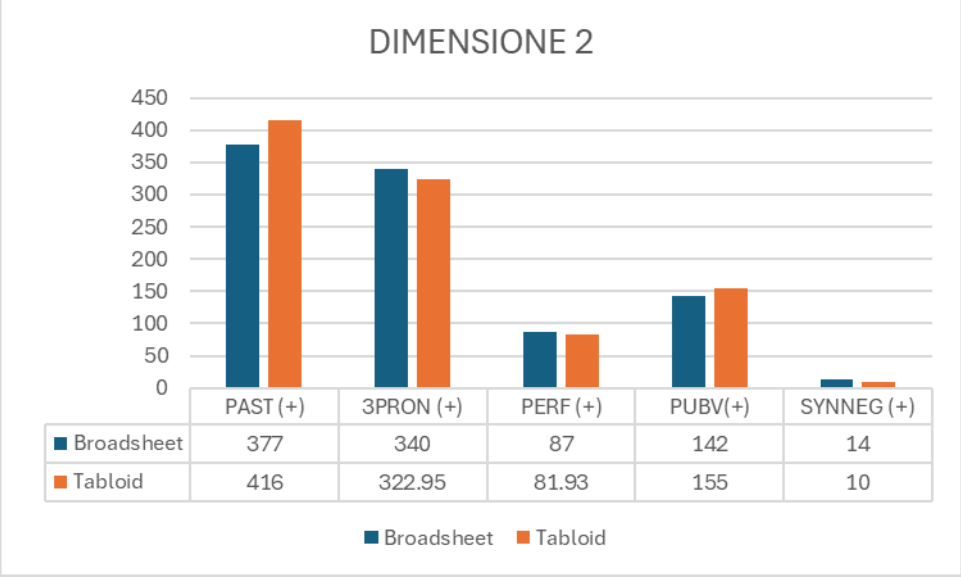
Nel corpus in analisi, il subcorpus dei *broadsheet* presenta un valore medio di *dimension score* di 2.11 e i *tabloid* di 2.07, con una differenza di 0.04. Entrambi i gruppi, tuttavia, si collocano in una fascia di debole narratività, indicando una prevalenza di contenuti non narrativi nei testi giornalistici relativi alle elezioni del leader del Partito Conservatore.

Figura 4



Le variabili linguistiche associate a un loading positivo su questa Dimensione - e dunque responsabili dell'incremento del valore narrativo di un testo - sono: tempo passato, pronomi di terza persona, aspetto perfettivo, verbi dichiarativi e negazione sintetica. Un testo che presenta frequenze elevate di queste caratteristiche tende a collocarsi più in alto lungo l'asse della narratività. Al contrario, un'alta frequenza di verbi al tempo presente e aggettivi attributivi marca il testo come meno funzionale alla narrazione.

Figura 5



Nel nostro caso, l'analisi delle frequenze relative ha confermato solo parzialmente l'orientamento più narrativo dei *broadsheet* rispetto ai *tabloid*. In particolare, tre delle cinque variabili sopra elencate - pronomi di terza persona (3PRON), aspetto perfettivo (PERF)e negazione sintetica (SYNNEG) - mostrano una frequenza più alta nel subcorpus dei *broadsheet*. Le altre due variabili, verbi al passato (PAST) e verbi pubblici (PUBV), non seguono questo pattern, introducendo un elemento di incoerenza rispetto all'atteso posizionamento stilistico.

Per investigare le cause di questa discrepanza, è stata condotta un'analisi supplementare attraverso gli strumenti offerti da *Sketch Engine*. In particolare, è stata generata una lista delle parole più frequenti filtrata sulla base della lista di **verbi dichiarativi** proposta nella guida del software *MAT*[9], utilizzando inoltre il filtro "find: verb". I risultati ottenuti da questa seconda analisi, più mirata, hanno prodotto i seguenti dati:

Tabella 6

GRUPPO	PUBV
Broadsheet	317.9
Tabloid	299.83

Affrontando l'analisi con un diverso strumento computazionale, i dati ottenuti attraverso *Sketch Engine* confermano - e in parte ribaltano - i risultati precedentemente forniti dal *MAT*. In particolare, per quanto riguarda i verbi dichiarativi, i *broadsheet* mostrano frequenze relative (FR) superiori rispetto ai *tabloid*, in contrasto con quanto osservato nei risultati iniziali, consultabili in "Appendice, *Medie Frequenze Relative variabili linguistiche per ciascun gruppo*".

L'analisi su Sketch Engine è stata implementata anche per i **verbi al passato** utilizzando l'espressione regolare *[tag="VVD"]* come *query* nello strumento *Concordance*. Questa ricerca, invece, non ha evidenziato discrepanze sostanziali rispetto ai risultati forniti dall'altro strumento adoperato in questo studio: pur differendo nei valori assoluti, le FR calcolate tramite *Sketch Engine*, visibili nella Tabella 7, confermano che i *tabloid* presentano una maggiore incidenza di verbi al passato, coerentemente con quanto emerso nell'analisi *MAT*.

Tabella 7

GRUPPO	PAST

[9] MAT v 1.3.2 - Manual.pdf - Google Drive, p.28; l'elenco di riferimento è espressamente lo stesso di Quirk et al. (1985: 1180–1).

Broadsheet	218.83
Tabloid	247.52

Tali incongruenze tra i due strumenti possono essere attribuite sia alle differenze metodologiche tra i sistemi di annotazione e conteggio, sia alla scarsa distanza tra i gruppi rispetto alla Dimensione 2, che risulta complessivamente debole per entrambi i subcorpora.

Per approfondire ulteriormente lo stile linguistico dei *broadsheet,* caratterizzati da un livello di narratività solo marginalmente superiore rispetto ai *tabloid*, si propone di seguito un estratto rappresentativo del subcorpus, utile a contestualizzare qualitativamente i dati quantitativi emersi:

The Daily Telegraph, file35719076
said conservatism must become a "team effort" with a "new respect" for members who she called the "backbone of communities". </s><s> Meanwhile, Mrs Braverman withdrew from the race, saying that her party does not want to hear the truth about why it lost the election. </s><s> She said there was no point in running "when most of the MPs disagree with my diagnosis", writing: "The traumatised party does not want to hear these things said out loud." </s><s> Mrs Braverman said she had the ten Tory MP

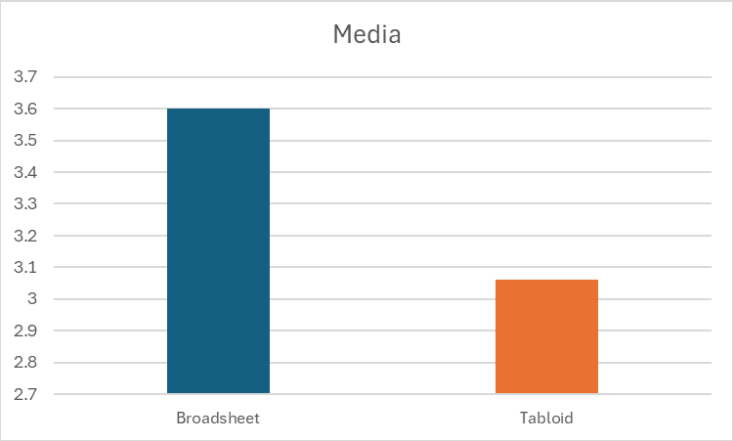
Il passo selezionato si distingue per l'elevata presenza di pronomi di terza persona e verbi pubblici - *say*, *write* e le rispettive forme flesse. In particolare, il verbo *say* risulta particolarmente frequente, occupando la terza posizione nella lista delle parole più frequenti complessiva relativa al subcorpus dei *broadsheet* (cfr. "Appendice, Lista dei primi 100 verbi più frequenti – *broadsheet*").

Analisi Dimensione 3

La terza Dimensione proposta da Biber consente di posizionare i testi lungo un continuum compreso tra discorso esplicito (*Explicit*) e discorso dipendente dal contesto (*Context-dependent*). Un elevato *dimension score* segnala un testo slegato dal contesto e quindi basato su riferimenti espliciti; al contrario testi con un basso punteggio saranno caratterizzati da riferimenti deittici e contestuali.

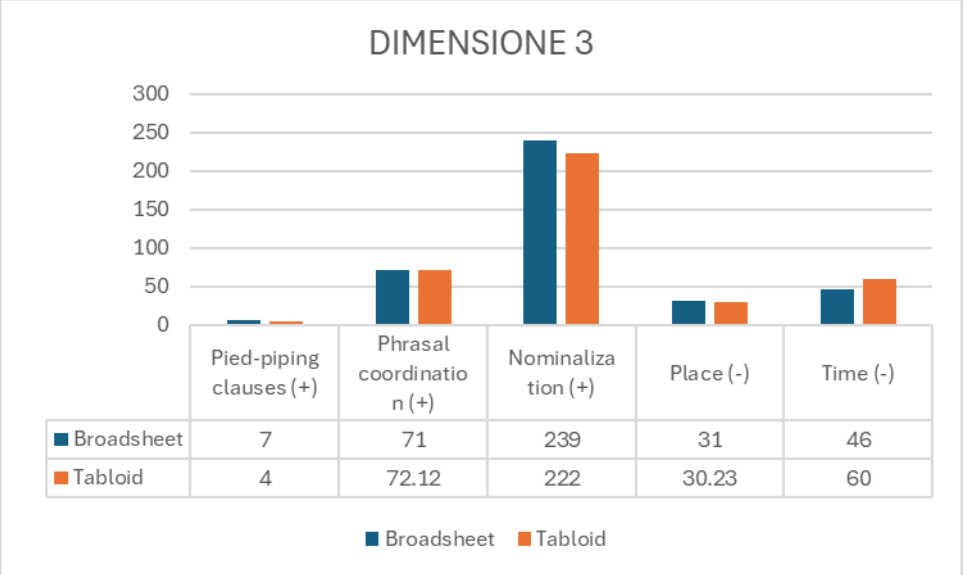
Nel presente studio, entrambi i subcorpora si collocano nel semiasse positivo, confermando una prevalente indipendenza dal contesto comunicativo, come prevedibile per testi di stampo giornalistico. I *broadsheet* raggiungono una media pari a 3.60, mentre i *tabloid* si attestano su 3.06, con una differenza tra i punteggi medi di 0.54. Sebbene modesta, tale distanza segnala una tendenza lievemente maggiore dei *broadsheet* a uno stile esplicito.

Figura 6



Questo dato è parzialmente corroborato dall'osservazione delle variabili con *factor loading* rilevante su questa Dimensione.

Figura 7



In particolare, i *broadsheet* registrano frequenze relative più elevate per le seguenti variabili: trascinamento sintattico (pied-piping) e nominalizzazione, e più ridotte per avverbi di tempo. Tuttavia, altre variabili chiave sembrano invertire il trend emerso dai *dimension score*s. In particolare, i *tabloid* presentano FR superiori per coordinazione di sintagmi e inferiore per avverbi di luogo.

Tale disallineamento tra i punteggi globali e la distribuzione delle singole variabili ha suggerito la necessità di una verifica incrociata. A questo scopo, sono sati ripetuti i calcoli relativi a nominalizzazione, coordinazione frasale, avverbi di tempo e avverbi di luogo utilizzando *Sketch Engine* come strumento alternativo al *MAT.*

I risultati ottenuti offrono una prospettiva utile a comprendere i motivi della discrepanza, in particolare, per quanto riguarda la coordinazione frasale. Analogamente, l'analisi della

nominalizzazione fornisce spunti rilevanti per comprendere la modesta distanza tra i due gruppi, distanza che può essere plausibilmente attribuita, almeno in parte, alla condivisione di un medesimo registro testuale - quello giornalistico -, caratterizzato da un'elevata coerenza stilistica e funzionale.

Un'analisi più approfondita dei dati ha evidenziato che il *Daily Express* contribuisce in modo significativo all'innalzamento della media delle frequenze relative (FR) della **coordinazione frasale** all'interno del gruppo *tabloid*. In assenza di questo quotidiano, la media del gruppo *tabloid* risulterebbe inferiore rispetto a quella osservata nel gruppo *broadsheet*.

Tabella 8

GRUPPO	Coordinazione frasale
Broadsheet	112.98
Tabloid	126.89
Daily Express	136.93
Tabloid-Daily Express	106.41

La raccolta dei dati è stata condotta tramite lo strumento *Concordanc*e della piattaforma *Sketch Engine*, applicando la seguente espressione regolare a tutti e quattro i gruppi considerati: *[tag="RB.?|J.*|V.*|N.*"] [tag="CC"] [tag="RB.?|J.*|V.*|N.*"]*. Questa formula consente di individuare strutture sintattiche caratterizzate da un elemento lessicale (avverbio, aggettivo, verbo o nome), seguito da una congiunzione coordinante e da un ulteriore elemento della stessa categoria. In seguito, sono state calcolate le frequenze relative per text type con lo strumento *Frequency* all'interno del subcorpus *tabloid* (Tabella 9). La densità relativa indica che la coordinazione tra sintagmi è particolarmente caratteristica del Daily Express, essendo 108 volte più frequente che nell'intero corpus.

Tabella 9

doc.newspaper	"Frequency"	"Relative density"	"Relative in text types"
Daily Express	924	107.91081	13693.14898
Daily Mail	164	76.94382	9763.64827
Daily Mirror	72	88.79597	11267.60563
The Sun	66	92.17126	11695.90643
Daily Star	50	92.75709	11770.24482

Anche per la *linguistic feature* della **nominalizzazione** è stato condotto un calcolo delle frequenze relative (FR), tramite lo strumento *Concordance* di *Sketch Engine*, impiegando la seguente espressione CQL: *[lemma=".*ity|ness|tion|ment"]*[10].

Tabella 10

GRUPPO	Nominalizzazione
Broadsheet	393,174
Tabloid	435,574
Daily Express	443,101
Tabloid-Daily Express	420,219

Questa seconda analisi, sebbene evidenzi un'inversione nei valori medi di FR tra i gruppi *tabloid* e *broadsheet* (con i *broadsheet* che presentano frequenze inferiori), risulta particolarmente significativa in quanto conferma, anche in questo caso, il ruolo centrale del *Daily Express* nella distribuzione della variabile linguistica analizzata.

Nella *Figura 8* viene presentato un istogramma in cui l'asse delle ascisse rappresenta la posizione della variabile nel subcorpus *tabloid*, mentre l'asse delle ordinate riporta la frequenza della *feature* linguistica in oggetto. Le barre che concentrano la porzione più consistente della frequenza complessiva - circa tra il 15% e il 32% dell'asse x - corrispondono a testi appartenenti al *Daily Express*.

Poiché i documenti sono ordinati per testata giornalistica è possibile affermare che anche in questo caso il *Daily Express* esercita un'influenza significativa sull'andamento della variabile nel subcorpus *tabloid*. Questo è confermato dallo strumento "*Frequency*" per "*text types*" applicato sempre alla stessa *query* e allo stesso subcorpus. I risultati sono riportati nella Tabella 11. La densità relativa superiore a 100% indica che la *query* cercata è caratteristica di quel giornale.

[10] Questa espressione restituisce tutti i lemmi terminanti per "-ity", "-ness", "-ment", o "-tion", senza evitare dunque parole che, pur contenendo le terminazioni elencate, non sono nominalizzazioni. Questo è il metodo adottato anche da MAT che per definire la variabile "Nominalizations" scrive: "Any noun ending in -tion, -ment, -ness, or -ity, plus the plural forms. Although Biber (1988) does not mention that this variables was checked manually, it is likely that a stop list was used to avoid obviously erroneous tagging (e.g. city). However, this was not indicated in the appendix of Biber (1988)." (MAT v 1.3.2 - Manual.pdf - Google Drive, p.19)

Figura 8

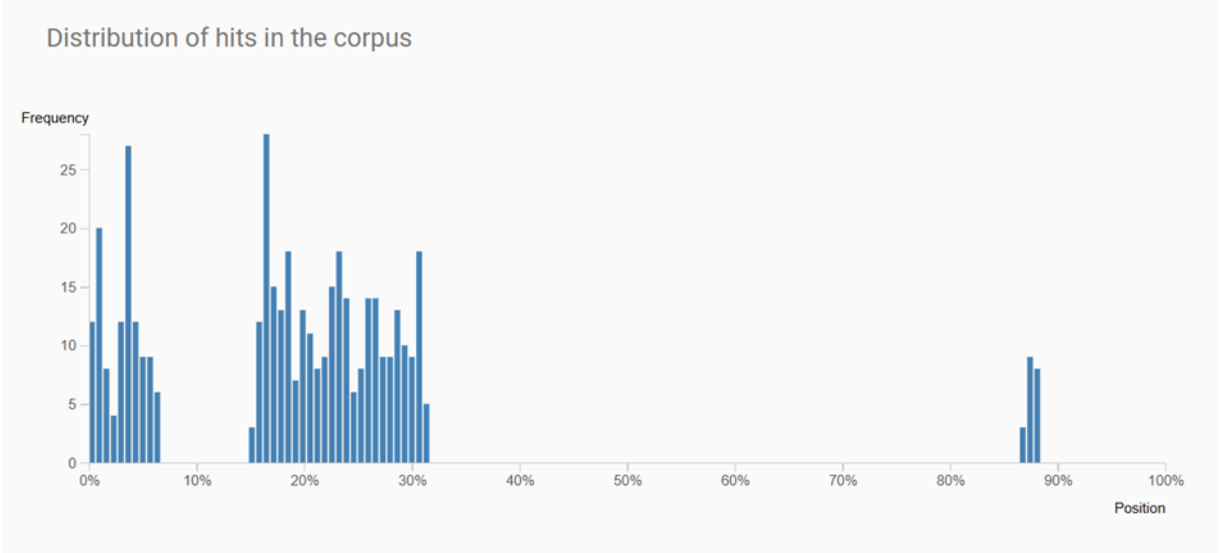


Tabella 11

doc.newspaper	"Frequency"	"Relative density"	"Relative in text types"
Daily Express	299	101.72806	4431.00817
The Daily Mail	75	102.51036	4465.08305
Daily Mirror	38	136.52775	5946.79186
The Sun	20	81.36884	3544.21407
Daily Star	6	32.42686	1412.42938

Per quanto riguarda, infine, le *feature* linguistiche relative agli **avverbi di tempo e luogo**, l'analisi condotta su *Sketch Engine* conferma la presenza di una differenza tra i due gruppi (*broadsheet* vs. *tabloid*), sebbene i valori rilevati non siano particolarmente marcati.

Tabella 12

GRUPPO	Place	Time
Broadsheet	8.05	28.32
Tabloid	7.46	30.13
Daily Express	7.71	30.38
Tabloid - Express	6.95	29.63

In sintesi, la variazione osservata tra i due sottogruppi può essere in parte attribuita alla distribuzione disomogenea delle occorrenze all'interno del subcorpus *tabloid*, influenzato in misura significativa

dalla presenza del *Daily Express*. Tale effetto è risultato evidente soprattutto in relazione alle variabili nominalizzazione e avverbi di tempo e luogo, con il potenziale di invertire le differenze tra i gruppi in esame.

Analisi Dimensione 4

La quarta Dimensione di Biber riguarda il grado di intenzionalità persuasiva espresso nel testo[11]. Consegue che testi con un punteggio elevato su questa Dimensione conterranno strutture linguistiche orientate all'espressione esplicita di opinioni e giudizi, con l'obiettivo di influenzare il destinatario. Al contrario, testi con un basso punteggio saranno caratterizzati da un linguaggio più neutro, distaccato e orientato all'oggettività.

Dall'analisi comparativa tra le medie dei *dimension score*s dei gruppi esaminati, emerge una considerevole differenza, di seguito illustrata dal grafico. Mentre i *broadsheet* registrano un punteggio medio di 2.76, i *tabloid* si attestano su una media di 1.49. Sebbene entrambi i valori non si scostino in modo marcato dallo 0, lo scarto tra le medie, pari a 1.27, marca una differenza significativa tra i due insiemi. Dal confronto si evince che i *broadsheet* tendono maggiormente verso l'estremo *Overt,* suggerendo una lieve ma percepibile inclinazione persuasiva nel loro linguaggio.

Figura 9

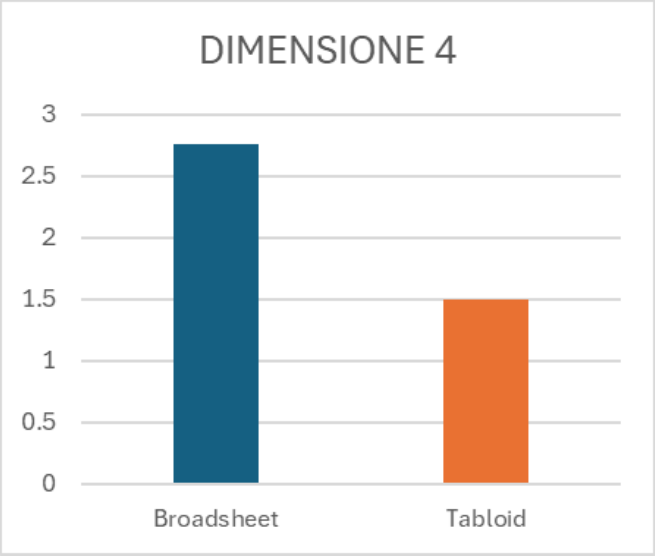


Tabella 13

GRUPPO	MEDIA DIMENSION SCORE	VALORE MINIMO	VALORE MASSIMO	STANDARD DEVIATION
BROADSHEET	2.758013468	-9.27	13.37	3.735509448

[11] Ivi, p. 111.

TABLOID	1.49225	-9.27	20.96	4.793410095

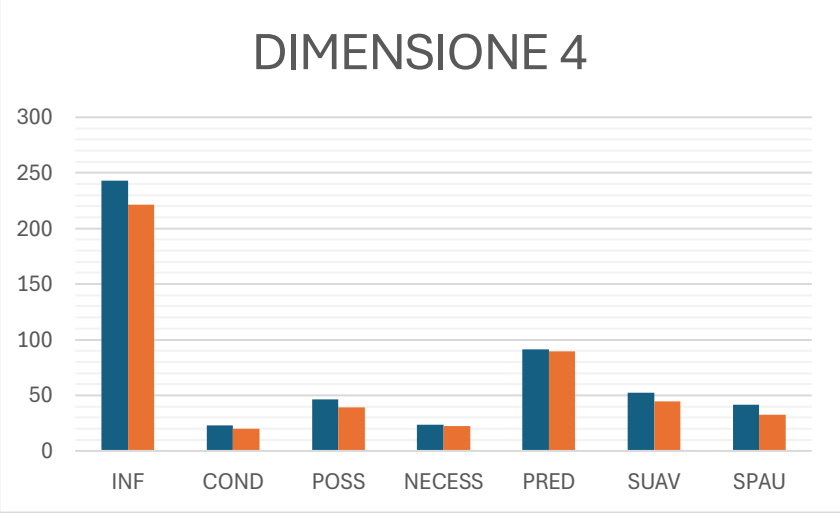
Per approfondire sul divario riscontrato, sono state comparate le medie delle frequenze relative, rilevate per ciascun subcorpus, delle variabili linguistiche associate a una marcata intenzione persuasiva, secondo la classificazione proposta da Biber[12]. Tra quest'ultime rientrano i verbi modali, le proposizioni infinitive e i verbi di persuasione[13].

Il confronto evidenzia scarti rilevanti, chiaramente visibili nella *Figura 10*, tra i due gruppi nei valori di tutte le variabili esaminate, confermando la presenza di una differenza tra gli insiemi.

Tabella 14

GRUPPO	INF	COND	POSS	NECESS	PRED	SUAV	SPAU
BROADSHEET	243.228	22.835	46.387	23.498	91.269	52.346	41.585
TABLOID	221.315	19.85	39.37	22.15	89.235	44.38	32.39

Figura 10



Per indagare ulteriormente sulle differenze riscontrate nell'uso delle variabili sotto esame, quest'ultime sono state rianalizzate utilizzando un altro programma, *Sketch Engine*.

Nello specifico, sono state generate tre espressioni regolari per il linguaggio *CQL* per lo strumento *Concordance* al fine di quantificare la presenza di proposizioni infinitive, verbi modali e verbi di persuasione.

Per identificare le **proposizioni infinitive**, denominate *INF nel* grafico, è stata applicata la seguente espressione:

[12] Ivi, p. 103.
[13] Ivi, Appendice II.

a) [lemma="to"][tag="VV"].

Il confronto tra le frequenze relative individuate mostra una differenza considerevole, che sembra confermare i risultati precedentemente ottenuti. I *broadsheet*, infatti, registrano un numero di frequenze relative – 16,471.51 – superiore a quello dei *tabloid,* con uno scarto pari a 102.68 per milione di token.

Figura 11 - Broadsheet

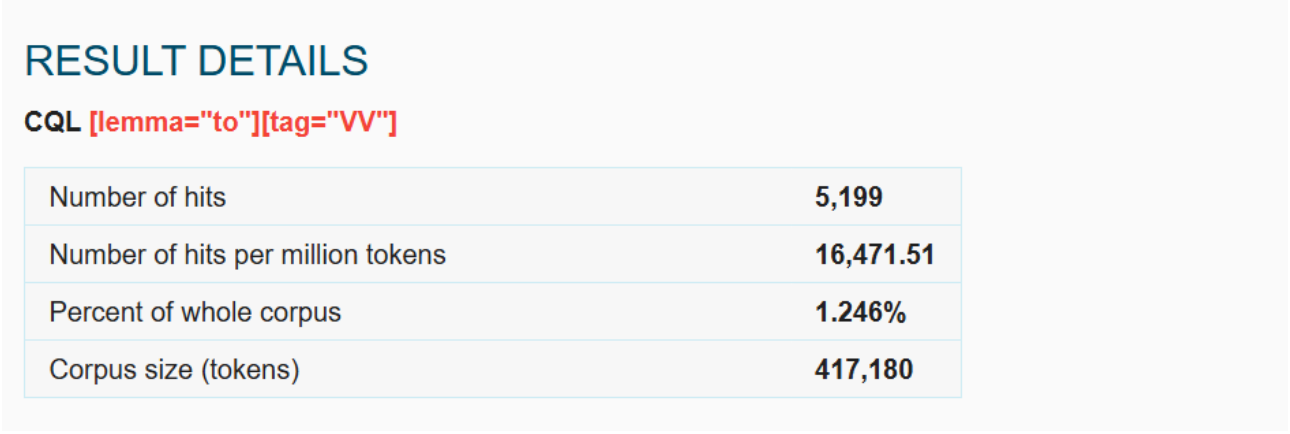


Figura 12 - Tabloid

Per l'individuazione dei **verbi modali** è stata usata la seguente espressione, che racchiude le categorie *POSS, PRED, NECESS*:

b) [tag="MD" & lemma="can|could|may|might|must|shall|should|will|would|ought"]

Figura 13 - Broadsheet

RESULT DETAILS

CQL [tag="MD" & lemma="can|could|may|might|must|shall|should|will|would|ought"]

Number of hits	**4,425**
Number of hits per million tokens	**14,019.31**
Percent of whole corpus	**1.061%**
Corpus size (tokens)	**417,180**

Figura 14 - Tabloid

RESULT DETAILS

CQL [tag="MD" & lemma="can|could|may|might|must|shall|should|will|would|ought"]

Number of hits	**1,405**
Number of hits per million tokens	**13,972.17**
Percent of whole corpus	**0.3368%**
Corpus size (tokens)	**417,180**

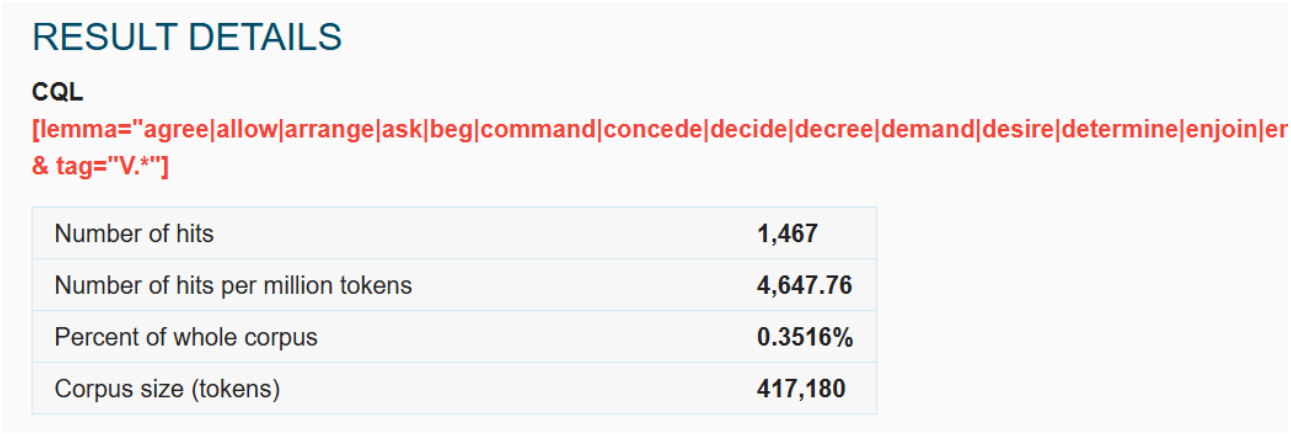
Dalla comparazione delle frequenze relative emerge uno scarto, seppure contenuto, pari a 47.14, suggerendo un uso maggiore dei verbi modali all'interno del subcorpus *broadsheet*. Quest'ultimi totalizzano, infatti, una frequenza media di 14,019.31, mentre i *tabloid* ottengono una media di 13,972.17.

Per la ricerca dei **verbi di persuasione**, ovvero *SAUV*, è stata usata la seguente espressione, volta a individuare i verbi dichiarativi definiti da Biber[14]:

c) [lemma="agree|allow|arrange|ask|beg|command|concede|decide|decree|demand|desire|determine|enjoin|ensure|entreat|grant|insist|instruct|intend|move|ordain|order|pledge|pray|prefer|pronounce|propose|recommend|request|require|resolve|rule|stipulate|suggest|urge|vote" & tag="V.*"]

[14] Ibidem.

Figura 15 - Broadsheet



RESULT DETAILS

CQL
[lemma="agree|allow|arrange|ask|beg|command|concede|decide|decree|demand|desire|determine|enjoin|er
& tag="V.*"]

Number of hits	1,467
Number of hits per million tokens	4,647.76
Percent of whole corpus	0.3516%
Corpus size (tokens)	417,180

Figura 16 – Tabloid

RESULT DETAILS

CQL
[lemma="agree|allow|arrange|ask|beg|command|concede|decide|decree|demand|desire|determine|enjoin|
& tag="V.*"]

Number of hits	419
Number of hits per million tokens	4,166.79
Percent of whole corpus	0.1004%
Corpus size (tokens)	417,180

Anche in questo caso, si registra una differenza a favore dei *broadsheet*, i quali registrano una frequenza media di 4,647.76. I *tabloid,* invece, si attestano su una media di 4,166.79, generando uno scarto significativo, pari a 480,97, conforme all'esito dell'analisi precedente.

Riguardo all'uso delle **particelle subordinanti** *if* e *unless,* rappresentate dalla variabile *COND* è stata usata l'espressione regolare al punto d):

d) [lemma="if|unless"]

Figura 17 - Broadsheet

RESULT DETAILS

CQL [lemma="if|unless"]

Number of hits	716
Number of hits per million tokens	2,268.44
Percent of whole corpus	0.1716%
Corpus size (tokens)	417,180

Figura 18 - Tabloid

RESULT DETAILS

CQL [lemma="if|unless"]

Number of hits	**189**
Number of hits per million tokens	**1,879.53**
Percent of whole corpus	**0.04530%**
Corpus size (tokens)	**417,180**

L'indagine ha confermato, ancora una volta, i risultati ottenuti dall'analisi effettuata su *MAT*, evidenziando una differenza rilevante tra i gruppi. I *broadsheet* sembrano usare maggiormente le particelle *if* e *unless*, totalizzando una frequenza media di 2,268.44 per milione di token. I *tabloid*, al contrario, con un punteggio medio di 1,879.53, generano uno scarto rilevante, pari a 388.91, mostrando una minore propensione nell'uso di tali elementi.

Complessivamente, per le ultime due variabili linguistiche esaminate, ovvero uso di verbi di persuasione e particelle subordinanti, si registra un divario maggiore.

Si evince che le analisi contrastive condotte su *Sketch Engine* confermano e rafforzano la validità i risultati emersi nel confronto precedente.

Il testo riportato di seguito, tratto dal quotidiano *The Independent*, contribuisce a evidenziare la presenza, seppur contenuta, all'interno della sottocategoria *broadsheet*, degli aspetti linguistici analizzati. L'analisi quantitativa della composizione lessicale del testo evidenzia, infatti, l'uso di proposizioni infinitive e verbi modali e la presenza della preposizione subordinante *if*.

The Independent, file35719076
He is hoping to draw a dividing line with Ms Badenoch, who has not promised to leave the ECHR *if elected, saying only that she would be willing to do so if necessary. </s><s> Mr Jenrick also said the Conservatives would stand for a "complete break from Labour's failing agenda" under his leadership.*

Analisi Dimensione 5

La quinta Dimensione di Biber distingue i testi in base al grado di astrazione del linguaggio. I testi collocati sul semiasse positivo tenderanno a esprimersi in modo più astratto e concettuale, mentre i testi posizionati sul semiasse negativo adotteranno un linguaggio più concreto e tangibile[15].

La comparazione delle frequenze relative delle medie dei *dimension scores* per ciascun gruppo evidenzia una tendenza opposta rispetto a questa funzione comunicativa. I *tabloid* si collocano sul semiasse negativo con un punteggio di -0.30, mentre i *broadsheet*, con un punteggio positivo di 0.30, mostrano una maggiore inclinazione verso l'astrazione. Di seguito la tabella con i valori rilevati per ciascun gruppo.

Figura 19

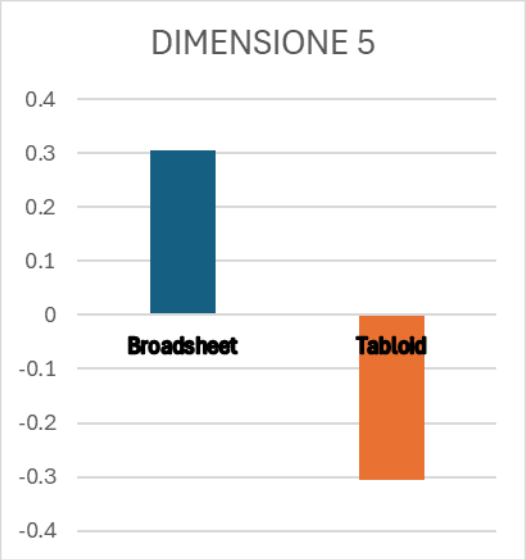


Tabella 15

GRUPPO	MEDIA DIMENSION SCORE	VALORE MINIMO	VALORE MASSIMO	DEVIAZIONE STANDARD
BROADSHEET	0.305925926	-3.92	13.53	2.319509334
TABLOID	-0.30615	-3.92	13.9	2.864217939

Sebbene i punteggi di entrambi i gruppi si collochino in prossimità dello 0, lo scarto tra i valori rappresenta una differenza significativa. Tale dissimilarità è stata approfondita mediante il confronto delle medie delle frequenze relative delle variabili linguistiche associate al linguaggio astratto.

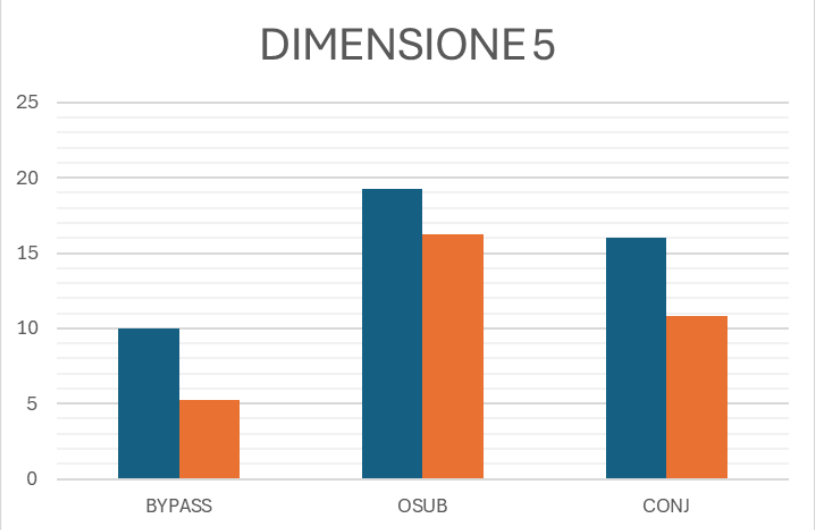
[15] Ivi, pp. 111-113.

In particolare, sono state esaminate le frequenze relative delle costruzioni passive introdotte dalla particella *by* (BYPASS), degli avverbi subordinativi (OSUB) e delle congiunzioni (CONJ), secondo la classificazione proposta da Biber[16]. Dal confronto emerge una marcata differenza tra i due gruppi, chiaramente rappresentata nel grafico sottostante.

Tabella 16

Figura 20

GIORNALE	BYPASS	OSUB	CONJ
BROADSHEET	9.996	19.265	16.04
TABLOID	5.22	16.22	10.8



Da una scrupolosa osservazione della *Tabella 16*, si nota uno scarto rilevante, pari a 4.77 nelle frequenze relative delle costruzioni passive seguite da *by* e un minore divario, seppur significativo, nelle occorrenze degli avverbi subordinativi, pari a 3.04. La differenza maggiore, pari a 5.24 viene registrata nel confronto tra i connettivi logici: mentre i *broadsheet* mostrano un numero di occorrenze superiore, pari a 16.04, i *tabloid*, con una media di 10.8, ottengono un punteggio più basso. Appare evidente che i *tabloid* mostrano un uso minore delle variabili linguistiche esaminate, suggerendo una

[16] Ibidem.

lieve propensione nell'uso di un linguaggio più concreto. Al contrario, i *broadsheet* sembrano mostrare una lieve ma percepibile inclinazione verso il linguaggio astratto.

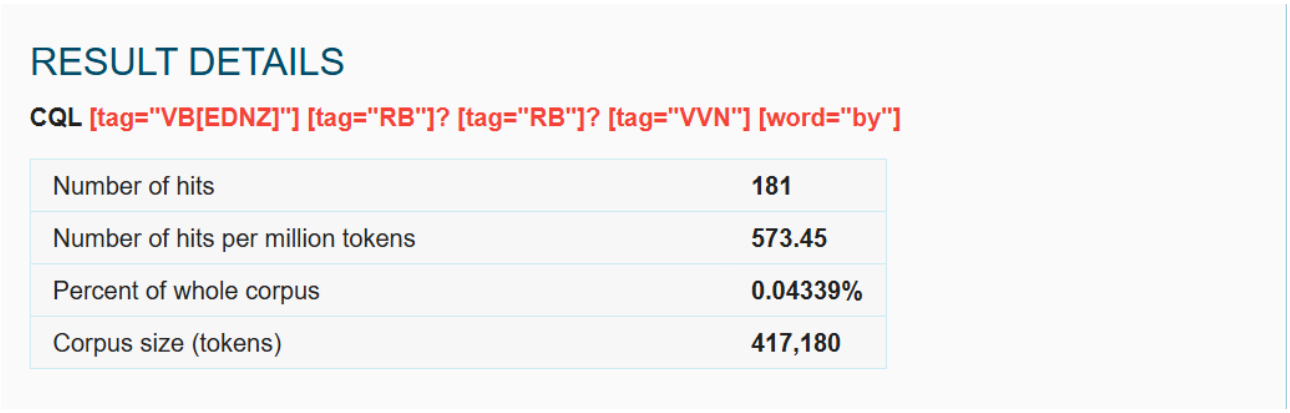
Al fine di verificare la robustezza dei i risultati ottenuti, sono state nuovamente esaminate le occorrenze delle variabili linguistiche sotto esame mediante *Sketch Engine*. Nello specifico, per l'individuazione delle variabili linguistiche è stata usato lo strumento *Concordance* e il linguaggio *CQL*.

Per quantificare la presenza della **costruzione passiva** seguita dalla particella *by* è stata applicata la seguente espressione regolare:

a) [tag="VB[EDNZ]"] [tag="RB"]? [tag="RB"]? [tag="VVN"] [word="by"]

Il confronto tra le occorrenze per milione di token restituisce una differenza significativa, con uno scarto superiore di 200 -FR per milione di token. Si registrano i seguenti valori: 573.45 per il gruppo *broadsheet* e 338.12 per i *tabloid*.

Figura 21 - Broadsheet

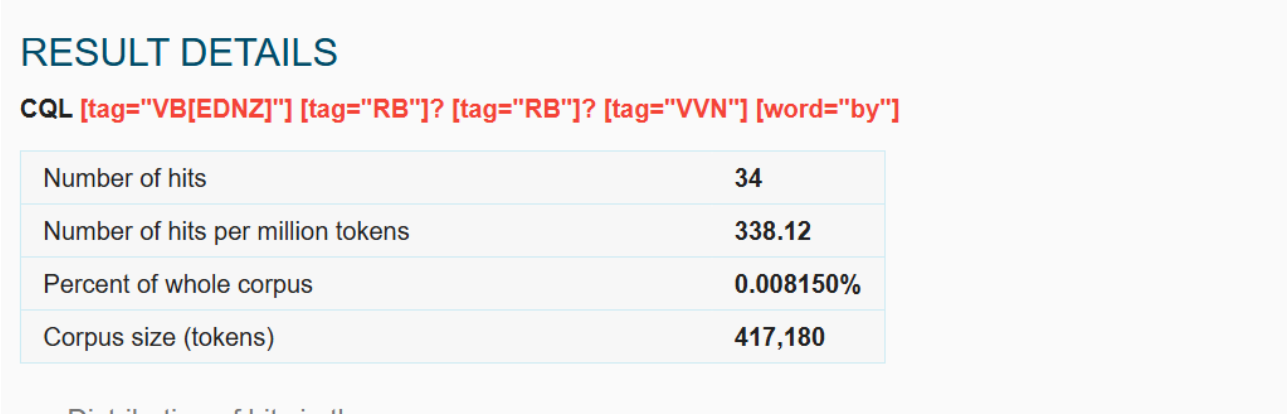


RESULT DETAILS

CQL [tag="VB[EDNZ]"] [tag="RB"]? [tag="RB"]? [tag="VVN"] [word="by"]

Number of hits	181
Number of hits per million tokens	573.45
Percent of whole corpus	0.04339%
Corpus size (tokens)	417,180

Figura 22 - Tabloid



RESULT DETAILS

CQL [tag="VB[EDNZ]"] [tag="RB"]? [tag="RB"]? [tag="VVN"] [word="by"]

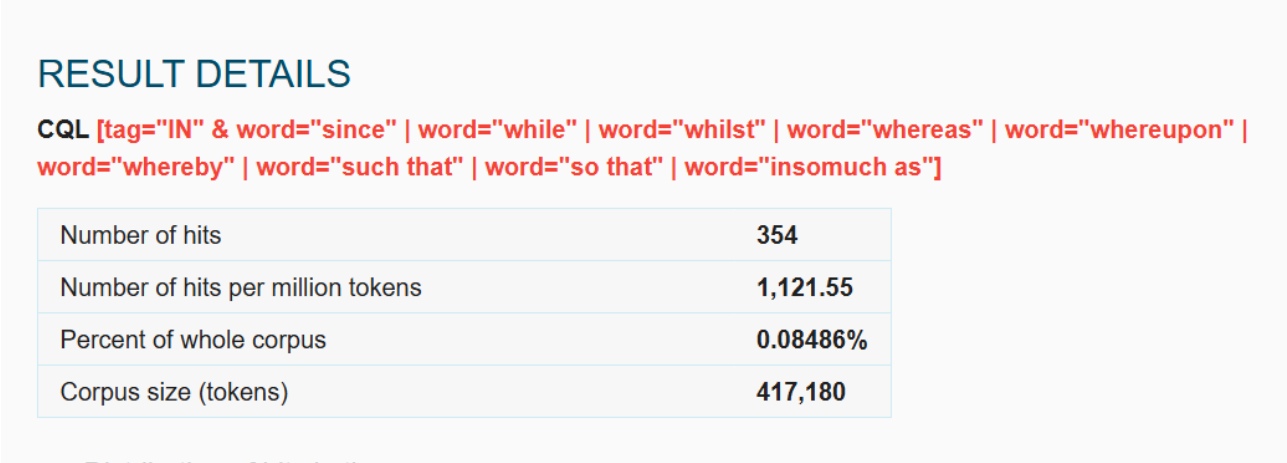
Number of hits	34
Number of hits per million tokens	338.12
Percent of whole corpus	0.008150%
Corpus size (tokens)	417,180

Con lo stesso approccio, è stata usata l'espressione regolare al punto *b)* per analizzare le frequenze relative degli **avverbi con funzione subordinativa**.

b) [tag="IN" & word="since" | word="while" | word="whilst" | word="whereas" | word="whereupon" | word="whereby" | word="such that" | word="so that" | word="insomuch as"]

I risultati mostrano un indice di frequenza superiore nei *broadsheet*, pari a 1,121.55 per milione di token; mentre i *tabloid* registrano una frequenza media significativamente più bassa, pari 1,014.35.

Figura 23 – Broadsheet

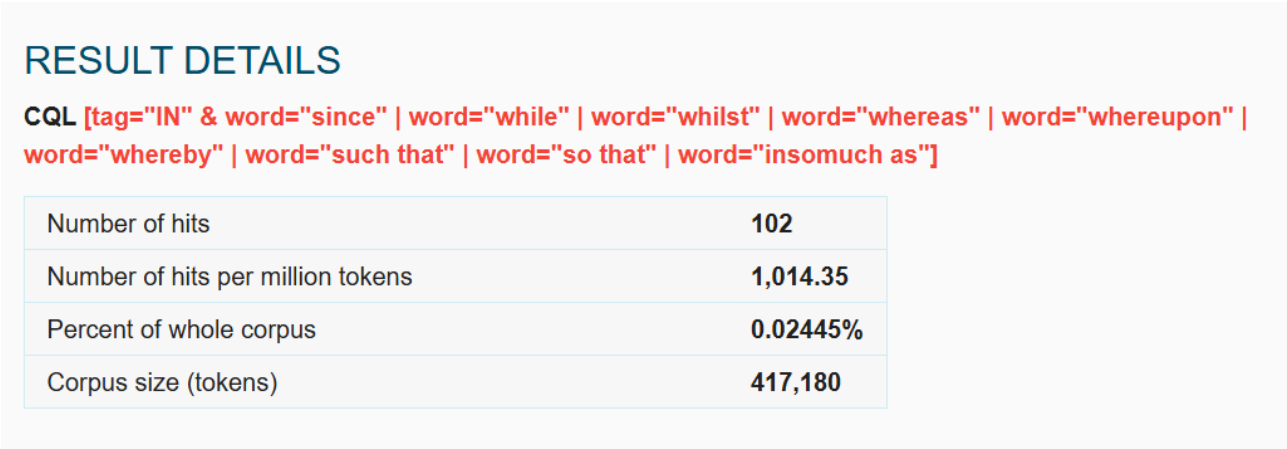


RESULT DETAILS

CQL [tag="IN" & word="since" | word="while" | word="whilst" | word="whereas" | word="whereupon" | word="whereby" | word="such that" | word="so that" | word="insomuch as"]

Number of hits	**354**
Number of hits per million tokens	**1,121.55**
Percent of whole corpus	**0.08486%**
Corpus size (tokens)	**417,180**

Figura 24 – Tabloid



RESULT DETAILS

CQL [tag="IN" & word="since" | word="while" | word="whilst" | word="whereas" | word="whereupon" | word="whereby" | word="such that" | word="so that" | word="insomuch as"]

Number of hits	**102**
Number of hits per million tokens	**1,014.35**
Percent of whole corpus	**0.02445%**
Corpus size (tokens)	**417,180**

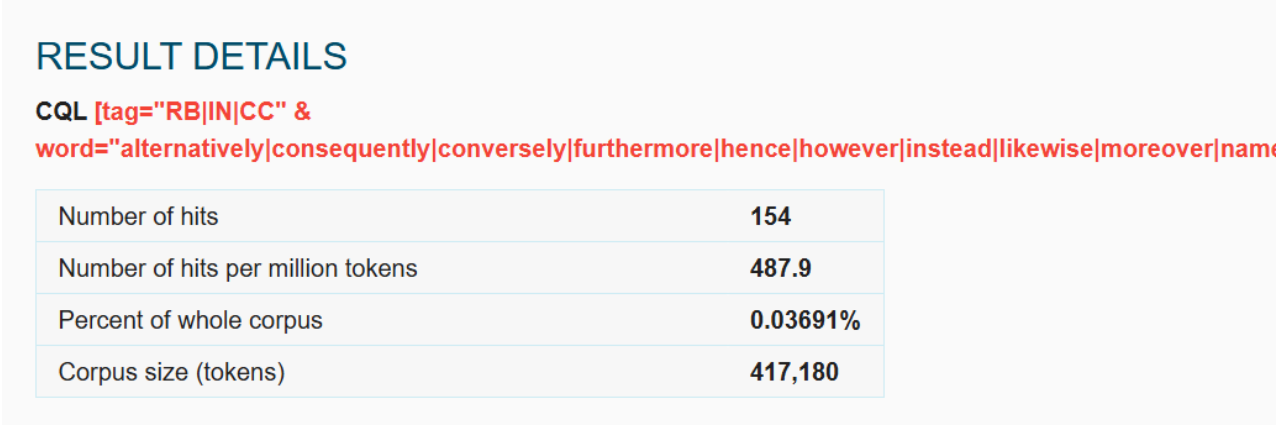
Con l'obiettivo di fornire un quadro completo, sono state analizzate anche le occorrenze di diversi **connettivi** tra quelli individuati da Biber[17], con la seguente espressione regolare:

c) [tag="RB|IN|CC" & word="alternatively|consequently|conversely|furthermore|hence|however|instead|likewise|moreover|namely|nevertheless|nonetheless|otherwise|notwithstanding|similarly|therefore|thus|viz"]

[17] Ivi, p. 236.

Anche in questo caso, l'analisi mette in evidenza una notevole differenza tra i gruppi nell'uso di tali elementi linguistici. I *broadsheet*, con un valore pari 487.9 per milione di token, ottengono un punteggio superiore a quello dei *tabloid*, i quali si attestano su una media di 367.95.

Figura 25 - Broadsheet

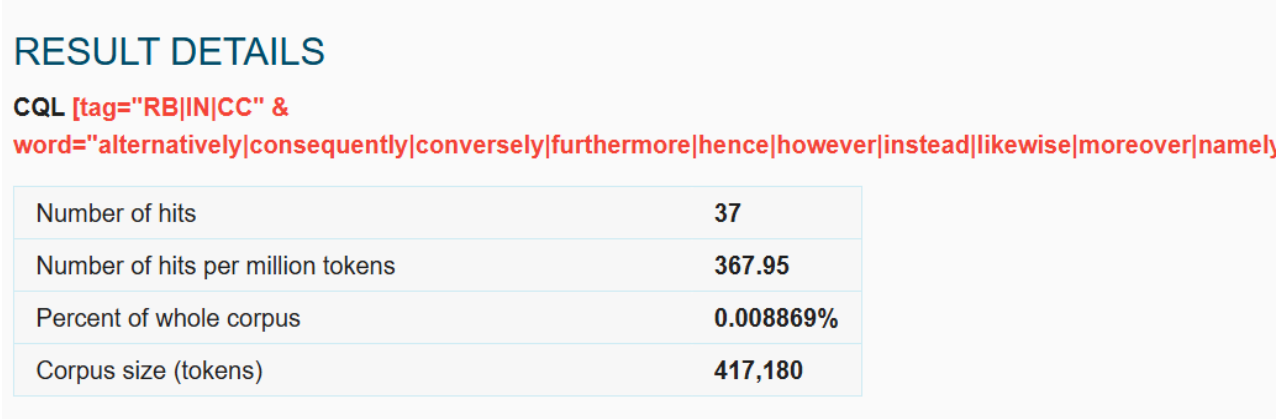


RESULT DETAILS

CQL [tag="RB|IN|CC" & word="alternatively|consequently|conversely|furthermore|hence|however|instead|likewise|moreover|name

Number of hits	**154**
Number of hits per million tokens	**487.9**
Percent of whole corpus	**0.03691%**
Corpus size (tokens)	**417,180**

Figura 26 - Tabloid



RESULT DETAILS

CQL [tag="RB|IN|CC" & word="alternatively|consequently|conversely|furthermore|hence|however|instead|likewise|moreover|namely

Number of hits	**37**
Number of hits per million tokens	**367.95**
Percent of whole corpus	**0.008869%**
Corpus size (tokens)	**417,180**

Complessivamente, tutte le analisi condotte su *Sketch Engine* confermano i risultati emersi dai confronti precedenti, mettendo in evidenza una netta differenza stilistica tra i due gruppi e una maggiore inclinazione dei *broadsheet* verso l'astrazione rispetto ai *tabloid*.

Di seguito l'estratto del *The Guardian*, tratto dal gruppo dei *broadsheet*, che illustra la presenza degli aspetti linguistici analizzati.

The Guardian, file35719076, doc241
CleverIy and Tugendhat, who are seen as centrists, are likely to mop up some of Stride's supporters. </s><s> Strategists on two other rival camps said Stride was suspected of privately working with Jenrick, however . </s><s> CleverIy's allies suggested some of the MPs who voted for him during the previous round may have switched to Tugendhat to ensure they both made it to the final four. </s><s> Some One Nation Conservatives are likely to argue that they should now unite behind either Tugendhat or CleverIy

Conclusioni

L'ipotesi iniziale di questo studio, ovvero l'esistenza di una **distinzione stilistica significativa** tra le testate giornalistiche *broadsheet* e *tabloid* nella copertura mediatica delle elezioni del leader del Partito Conservatore britannico, ha trovato parziale conferma nell'analisi condotta. L'applicazione dell'*Hierarchical Agglomerative Cluster Analysis* ha infatti evidenziato la formazione di due cluster distinti che rispecchiano, in larga misura, la classificazione preliminare delle testate. Tuttavia, si segnala un'eccezione rilevante: il *Daily Express*, pur appartenente al gruppo *tabloid*, si aggrega al cluster dei *broadsheet*, suggerendo una maggiore affinità stilistica con quest'ultimo gruppo.

Rimane tuttavia necessario un approfondimento delle dimensioni specifiche lungo le quali si articola la variazione stilistica tra i gruppi individuati. I risultati della *Multidimensional Analysis* mostrano una differenziazione contenuta tra i *dimension scores* dei due gruppi, con gli scarti più significativi registrati per le dimensioni 4 e 5, pari rispettivamente a 1.27 e 0.62. La **Dimensione 4**, associata alla *tendenza persuasiva*, evidenzia valori prossimi allo zero per entrambi i gruppi, ma mostra comunque una separazione considerevole tra di essi. La **Dimensione 5**, legata al grado di *astrazione*, si rivela invece particolarmente esplicativa: i *tabloid* presentano valori negativi, indicando una minore astrattezza, mentre i *broadsheet* si posizionano su valori positivi, coerenti con una maggiore tendenza all'astrazione. Questo risultato è in linea con le previsioni teoriche che associano uno stile più concreto e accessibile ai *tabloid*.

Le altre dimensioni mostrano variazioni meno marcate:

- La Dimensione 1 (*Informational vs. Involved*) evidenzia uno scarto pari a 0.51, con i *tabloid* orientati maggiormente verso l'estremo *Informational* rispetto ai *broadsheet*.
- La Dimensione 2 (*Narrative vs. Non-Narrative*) restituisce valori pari a 2.11 per i *broadsheet* e 2.07 per i *tabloid*, indicando una tendenza leggermente maggiore alla *narrazione* nei *broadsheet* (scarto: 0.04).
- La Dimensione 3 (*Explicit vs. Situation-Independent*) attribuisce ai *broadsheet* un punteggio di 3.60 e ai *tabloid* di 3.06, con una differenza reciproca di 0.54, connotando quindi i *broadsheet* come più espliciti e i *tabloid* come più simili (anche se in termini molto ridotti) ai testi conversazionali.

A supporto dell'analisi multidimensionale, è stato condotto un confronto tra le frequenze relative delle variabili linguistiche con *factor loading* superiore a 0.35. I risultati confermano solo parzialmente i posizionamenti dei gruppi sulle varie dimensioni: in alcuni casi le frequenze relative supportano i *dimension scores*, in altri se ne discostano. Questa apparente incongruenza non inficia necessariamente la validità dello studio, ma va contestualizzata sulla base di alcune considerazioni:

- I due gruppi - *broadsheet e tabloid* - condividono l'appartenenza a un **medesimo registro** comunicativo, quello dei *press editorials*, che implica una relativa omogeneità stilistica di fondo. Pertanto, anche differenze minime nei punteggi - inferiori a 1 - possono risultare statisticamente o stilisticamente rilevanti.
- L'influenza del ***Daily Express*** emerge chiaramente nella *cluster analysis* e contribuisce ad attenuare le differenze tra i due sottogruppi, come osservato nell'interpretazione della Dimensione 3.
- È stata osservata un'elevata **variabilità interna ai subcorpora**, come evidenziato dalle deviazioni standard prossimi o superiori ai valori medi dei *dimension scores*. Una lieve riduzione della variabilità si osserva nei dati relativi ai *broadsheet*, che presentano *Coefficient of Variation* inferiori rispetto al corpus generale per tutte le dimensioni analizzate, eccetto la 5 e la 6. Per la consultazione dei valori di *standard deviation* e *coefficient of variation* si rimanda all'Appendice, *Confronti Dimension Scores*.

Inoltre, per garantire una maggiore robustezza metodologica, alcune misurazioni sono state replicate utilizzando strumenti diversi. In particolare, *Sketch Engine* ha in alcuni casi prodotto risultati discordanti rispetto ai dati statistici ottenuti con *MAT*, a dimostrazione di come la scelta dello strumento influenzi le evidenze empiriche e debba quindi essere considerata attentamente in fase interpretativa.

Nel complesso, la dicotomia *tabloid/broadsheet* emerge come una categoria interpretativa utile per analizzare la variazione stilistica nel registro dei *press editorials*. Infatti, dai *coefficient of variation* è possibile dedurre che la categoria dei *broadsheet* costituisce all'interno del corpus un gruppo compatto e stilisticamente omogeneo, in particolare lungo le dimensioni 1-4. Al contrario, il gruppo dei *tabloid* registra CV di gran lunga superiori, sia ai *broadsheet* che al corpus intero. Simili risultati suggeriscono che questa distinzione non è pienamente esaustiva: ulteriori fattori, quali l'identità editoriale di singole testate (es. *Daily Express*) o l'orientamento politico dei giornali, potrebbero offrire una classificazione più fine e aderente alle reali dinamiche stilistiche. In particolare, un'analisi specificamente centrata sul *Daily Express* potrebbe consentire di isolare pattern linguistici distintivi, mentre una classificazione delle testate in base all'orientamento politico (destra/sinistra) – esclusa nella fase preliminare per ragioni di economia analitica – si configura come un'opzione promettente per ulteriori possibili sviluppi della ricerca.

Appendice

Variabili linguistiche

1. PAST
2. PERF
3. PRES
4. PLACE
5. TIME
6. 1PRON
7. 2PRON
8. 3PRON
9. IT
10. DEMPRON
11. INDPRON
12. DO
13. WHQU
14. NOMZ
15. GER
16. NN
17. PASS
18. BYPASS
19. BE
20. EXIST
21. THVC
22. THAC
23. WHCL
24. INF
25. PRESP
26. PASTP
27. WZPAST
28. WZPRES
29. TSUB
30. TOBJ
31. WHSUB
32. WHOBJ
33. PIRE
34. SERE
35. CAUS
36. CONC
37. COND
38. OSUB
39. PP
40. JJATR
41. JJPRED
42. ADV
43. TTR
44. MWL
45. CONJ
46. DWNT
47. HDG
48. AMP
49. EMPH
50. DPAR
51. DEMO
52. POSS
53. NECESS
54. PRED
55. PUBV
56. PRIV
57. SUAV
58. SMAP
59. CONT
60. THATD
61. STPR
62. SPINF
63. SPAU
64. PHC
65. ANDC
66. SYNEG
67. ANNEG

MAT Analysis – Frequenze relative variabili linguistiche con Stanford Tagging

Consultabile al seguente link: https://docs.google.com/spreadsheets/d/19DI_qU6yH5_epizeJMdOPD90d88Y3vnWjFtTYpFN__Y/edit?usp=drive_link

Frequenze relative variabili linguistiche normalizzate – Biber comparison

Consultabile al seguente link: https://docs.google.com/spreadsheets/d/14xLIwLwYnXl--eawV8lyLrQdQcBuimai/edit?usp=sharing&ouid=102264050224175720794&rtpof=true&sd=true

Legenda:

Tabella 17

BN	I
BN_B	The Independent
BL_A	The Guardian
BR_A	The Times
BR_B	The Daily Telegraph
TN_A	Daily Star
TL_A	Daily Mirror
TR_A	The Sun
TR_B	Daily Mail
TR_C	Daily Express

La prima lettera delle etichette indica se il giornale è un *broadsheet* o un *tabloid*, la seconda se è di sinistra (L), destra (R) o neutro/di centro (N), la terza serve a fare distinzione tra giornali che condividono la prima coppia di lettere (es. The Times e The Daily Telegraph sono entrambi *broadsheet* di destra).

Medie Frequenze Relative variabili linguistiche per ciascun gruppo

Le medie delle FR delle variabili linguistiche individuate da Biber calcolate per ciascun gruppo sono consultabili al seguente link: https://docs.google.com/spreadsheets/d/1K61OywJKvj0unSXAETOS-y-mT54uZHxg/edit?usp=sharing&ouid=102264050224175720794&rtpof=true&sd=true

MAT Analysis – Dimension scores

Il dataset generato da MAT contenente i *dimension scores* calcolati per ogni testo è consultabile al seguente link: https://docs.google.com/spreadsheets/d/1sSd4IQ9b4E_P2I_m2Esz8-85KEvHpicPipJc4X2L9_I/edit?usp=sharing

Confronti Dimension Scores

Tabella 18

Corpus	DIM1	DIM2	DIM3	DIM4	DIM5	DIM6
media	-10.13	2.09	3.38	2.25	0.06	-1.13
min	-27.37	-6.97	-5.45	-9.27	-3.92	-3.50
max	22.95	21.22	14.51	20.96	13.90	5.11
SD*	7.18	3.31	3.11	4.23	2.57	1.49
CV	32.24%	14.86%	13.96%	18.99%	11.54%	6.69%

Tabella 19

Broadsheet	DIM1	DIM2	DIM3	DIM4	DIM5	DIM6
media	-9.92	2.11	3.60	2.76	0.31	-1.00
min	-23.71	-6.97	-3.27	-9.27	-3.92	-3.50
max	12.75	21.22	14.51	13.37	13.53	3.67
SD*	6.65	3.01	2.78	3.74	2.32	1.35
CV	3.89%	8.31%	4.48%	7.91%	43.49%	7.85%

Tabella 20

Tabloid	DIM1	DIM2	DIM3	DIM4	DIM5	DIM6
media	-10.43	2.07	3.06	1.49	-0.31	-1.32
min	-27.37	-5.93	-5.45	-9.27	-3.92	-3.50
max	22.95	16.41	12.96	20.96	13.90	5.11
SD*	7.88	3.70	3.52	4.78	2.86	1.65
CV	55.85%	26.22%	24.95%	33.88%	20.26%	11.69%

*Le *standard deviation* sono state ottenute con la funzione STDEV di *GoogleSheet*

Dataset per Cluster Analysis

Consultabile al seguente link: https://docs.google.com/spreadsheets/d/10V7bBkLyxHp9HpyhlA-hoHxB4A-y4M5_kFEgU7pnRl8/edit?usp=sharing

Legenda:

Tabella 21

Registro	Testate giornalistiche - Broadsheet
News_reportage	The Guardian
News_editorials	I
News_reviews	The Independent
Religion	The Times
Skill	The Daily Telegraph

Tabella 22

Registro	Testate giornalistiche - Tabloid
Popular_lore	Daily Mirror
Biography	Daily Star
Misc_government	The Sun
Academic	Daily Mail
Fiction_general	Daily Express

Lista dei primi 100 verbi più frequenti - broadsheet

Tabella 23

Item	Frequency	Relative frequency
be	11276	35724.69554
have	4159	13176.57048
say	2586	8192.98179
do	1564	4955.07483
make	822	2604.26567
get	657	2081.51161
go	654	2072.007
take	571	1809.04586
need	540	1710.83146
win	526	1666.47657
think	523	1656.97196
want	469	1485.88881
see	455	1441.53392
come	445	1409.85185
tell	382	1210.25485
give	335	1061.34915
leave	332	1051.84453
become	310	982.14399
work	287	909.27524
vote	269	852.24753
know	268	849.07932
lead	245	776.21057
ask	240	760.36954
lose	237	750.86492
put	210	665.32335
believe	207	655.81873
look	203	643.1459
add	202	639.9777
suggest	194	614.63204
try	189	598.79101
stand	187	592.4546
call	187	592.4546
mean	186	589.28639
set	183	579.78177
speak	167	529.09047
seem	165	522.75406
run	161	510.08123
claim	160	506.91303
change	156	494.2402
announce	156	494.2402
face	155	491.07199

use	155	491.07199
choose	154	487.90379
include	150	475.23096
show	149	472.06276
remain	147	465.72634
back	146	462.55814
talk	143	453.05352
find	138	437.21249
deliver	135	427.70787
keep	133	421.37145
expect	132	418.20325
bring	131	415.03504
unite	131	415.03504
start	124	392.8576
fail	124	392.8576
appear	119	377.01656
hold	117	370.68015
turn	114	361.17553
help	114	361.17553
happen	114	361.17553
elect	113	358.00732
feel	113	358.00732
serve	111	351.67091
support	111	351.67091
describe	110	348.50271
decide	109	345.3345
let	107	338.99809
argue	103	326.32526
end	103	326.32526
understand	101	319.98885
begin	100	316.82064
cut	99	313.65244
spend	98	310.48423
like	97	307.31602
build	97	307.31602
follow	94	297.8114
stop	92	291.47499
pay	91	288.30678
focus	89	281.97037
write	88	278.80216
replace	85	269.29755
allow	85	269.29755
agree	84	266.12934
eliminate	83	262.96113
accuse	81	256.62472

declare	80	253.45651
warn	80	253.45651
pick	80	253.45651
promise	79	250.28831
fight	76	240.78369
play	75	237.61548
launch	75	237.61548
offer	74	234.44727
seek	73	231.27907
receive	72	228.11086